R Notebook

```
# import data
data1<-read.table("/Users/phamkhoa/Documents/university/3/stats_model_1/w5/canopycover.txt", sep="\t", 
attach(data1)
```

```
#p1
library(betareg)
#a
a.norm<-glm(canopycover~factor(species) + basalarea + dbh.mean, family = gaussian("identity"), data=data
a.gamma<-glm(canopycover~factor(species) + basalarea + dbh.mean, family = Gamma("log"), data=data1)
a.ig<-glm(canopycover~factor(species) + basalarea + dbh.mean, family =inverse.gaussian("log"), data=data
a.beta<-betareg(canopycover~factor(species) + basalarea + dbh.mean, data=data1, link=c("logit"))

AIC(a.norm)
```

```
## [1] -275.7166
```

```
AIC(a.gamma)
```

```
## [1] -372.2039
```

```
AIC(a.ig)
```

```
## [1] -300.5095
```

```
AIC(a.beta)
```

```
## [1] -408.0153
```

```
#Choose beta link model
```

```
#b
beta.main<-betareg(canopycover~ basalarea + dbh.mean+factor(species), data=data1, link=c("logit"))
newdata<-data.frame(basalarea = 20, dbh.mean = 15, species = "pine")
pred<-predict(beta.main, newdata = newdata, type = "response")
pred
```

```
##         1
## 0.1953721
```

```
#c
eta<-predict(beta.main, newdata = newdata, type = "link")
xf <- c(0,20,15,1)
cov.eta<-t(xf)%*%vcov(beta.main)[-5,-5]%*%xf

lowerbound<-exp(eta-qnorm(0.975)*sqrt(cov.eta))/(1+exp(eta-qnorm(0.975)*sqrt(cov.eta)))
upperbound<-exp(eta+qnorm(0.975)*sqrt(cov.eta))/(1+exp(eta+qnorm(0.975)*sqrt(cov.eta)))
lowerbound
```

```
##           [,1]
## [1,] 0.1716804
```

```
upperbound
```

```
##           [,1]
## [1,] 0.2214591
```

```
#d
library(mvtnorm)
eta.f<-predict(beta.main, newdata = newdata, type = "link")
phi.hat<-coef(beta.main)[5]
xf <- c(0,20,15,1)
etarow<-c(t(xf),0)
phirow<-c(rep(0,length(xf)),1)
A<-rbind(etarow,phirow)

cov.etaphi<-A%*%vcov(beta.main)%*%t(A)
etaphi.star<-rmvnorm(1000, mean = c(eta.f,phi.hat), sigma = cov.etaphi)

muf.star<-exp(etaphi.star[,1])/(1+exp(etaphi.star[,1]))
phi.star<-etaphi.star[,2]

p.star<-muf.star*phi.star
q.star<-phi.star*(1-muf.star)

yf.star<-rbeta(1000, shape1=p.star, shape2=q.star)

lower.bound<-quantile(yf.star, c(0.1))
upper.bound<-quantile(yf.star, 1-c(0.1))
lower.bound
```

```
##       10%
## 0.1453159
```

```
upper.bound
```

```
##       90%
## 0.2528612
```

```
#e
library(lmtest)
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
beta.2way<-betareg(canopycover~basalarea+dbh.mean + factor(species) + basalarea:factor(species) + dbh.me
beta.H0<-betareg(canopycover~basalarea+ factor(species)+ basalarea:factor(species), data=data1, link=c(

lrtest(beta.H0, beta.2way)
```

```
## Likelihood ratio test
##
## Model 1: canopycover ~ basalarea + factor(species) + basalarea:factor(species)
## Model 2: canopycover ~ basalarea + dbh.mean + factor(species) + basalarea:factor(species) +
##     dbh.mean:factor(species)
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1   5 166.99
## 2   7 248.28  2 162.58  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lrtest(beta.H0, beta.2way)$Chisq[2]
```

```
## [1] 162.5798
```

```
#p2 import data
data2<-read.table("/Users/phamkhoa/Documents/university/3/stats_model_1/w5/NitrogenYield.txt", sep="\t"
attach(data2)
```

```
#a
a.poly<-lm(Yield~Nitrogen + I(Nitrogen^2))
coef(a.poly)[3]
```

```
## I(Nitrogen^2)
##  -0.002223739
```

```
#b
b<-glm(Yield ~ log(Nitrogen), family = gaussian("log"), data = data2)
newdata <- data.frame(Nitrogen = 150)
predict(b, newdata = newdata, type = "response")
```

```
##        1
## 90.66521
```

```
#c
library(nlme)
c<-nls(Yield~SSasymp(Nitrogen, Asym,R0,lrc), data=data2)
coef(c)[1]
```

```
##     Asym
## 102.0634
```

```
#d
d<-nls(Yield~SSmicmen(Nitrogen, Vm, K), data=data2)
predict(d, newdata = newdata, type = "response")
```

```
## [1] 90.57802
## attr(,"gradient")
##             Vm          K
## [1,] 0.7750935 -0.4680429
```

```
#e
xf<-c(150)
beta<-coef(c)
cov.beta<-vcov(c)
sigma<-sigma(c)

library(mvtnorm)
beta.star<-rmvnorm(1000, mean = beta, sigma = cov.beta)

Asym<-beta.star[,1]
R0<-beta.star[,2]
lrc<-beta.star[,3]

mu.star<-Asym+(R0-Asym)*exp(-exp(lrc)*newdata$Nitrogen)

yf.star<-rnorm(1000, mean = mu.star, sd =sigma)
pred.lowerbound<-quantile(yf.star, c(0.1))
pred.upperbound<-quantile(yf.star, c(0.9))
pred.lowerbound
```

```
##      10%
## 83.00629
```

```
pred.upperbound
```

```
##      90%
## 98.89287
```

```
#p3 import data
data3<-read.table("/Users/phamkhoa/Documents/university/3/stats_model_1/w5/caffeine.txt", sep="\t", dec=
attach(data3)
```

```
#a
# First, we test with the main model
# from the course,
# since poission distribution is for count data model
# bernuolli and multinomial logit is for catergorical data model
# beta require Y in a specific range
# distribution that suitable for this type of data is
# gaussian, inverse gaussian, gamma
# in each distribution, we use default link

p3a.gaussian<-glm(Caffeine~factor(Brand)+factor(Formulation), family = gaussian("identity"), data = data
p3a.inversegaussian<-glm(Caffeine~factor(Brand)+factor(Formulation), family = inverse.gaussian("1/mu^2")
p3a.gamma<-glm(Caffeine~factor(Brand)+factor(Formulation), family = Gamma("inverse"), data = data3)

# We use the AIC score, MSE, Shapiro to test
#AIC
AIC(p3a.gaussian)
```

```
## [1] 1427.895
```

```
AIC(p3a.inversegaussian)
```

```
## [1] 1418.676
```

```
AIC(p3a.gamma)
```

```
## [1] 1420.002
```

```
# MSE
mean(residuals(p3a.gaussian, type="response")^2)
```

```
## [1] 21.7226
```

```
mean(residuals(p3a.inversegaussian, type="response")^2)
```

```
## [1] 20.49858
```

```
mean(residuals(p3a.gamma, type="response")^2)
```

```
## [1] 20.88572
```

```
#Shapiro test
shapiro.test(residuals(p3a.gaussian, type="pearson"))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(p3a.gaussian, type = "pearson")
## W = 0.99122, p-value = 0.1592
```

```
shapiro.test(residuals(p3a.inversegaussian, type="pearson"))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(p3a.inversegaussian, type = "pearson")
## W = 0.99389, p-value = 0.4384
```

```
shapiro.test(residuals(p3a.gamma, type="pearson"))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(p3a.gamma, type = "pearson")
## W = 0.99346, p-value = 0.3766
```

```
#So, we will choose the inverse gaussian model with canonical link
# Now, we must test the variable
ig.1<-glm(Caffeine~factor(Brand), family = inverse.gaussian("1/mu^2"), data = data3)
ig.2<-glm(Caffeine~factor(Formulation), family = inverse.gaussian("1/mu^2"), data = data3)
ig.main<-glm(Caffeine~factor(Brand)+factor(Formulation), family = inverse.gaussian("1/mu^2"), data = da
ig.full<-glm(Caffeine~factor(Brand)*factor(Formulation), family = inverse.gaussian("1/mu^2"), data = da

#using l
library(lmtest)
lrtest(ig.main, ig.full)
```

```
## Likelihood ratio test
##
## Model 1: Caffeine ~ factor(Brand) + factor(Formulation)
## Model 2: Caffeine ~ factor(Brand) * factor(Formulation)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   4 -705.34
## 2   5 -686.02  1 38.645  5.085e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
AIC(ig.full)
```

```
## [1] 1382.031
```

```
AIC(ig.main)
```

```
## [1] 1418.676
```

```
AIC(ig.1)
```

```
## [1] 1446.657
```

```
AIC(ig.2)
```

```
## [1] 1471.966
```

```
# So, based on anova test and AIC, we choose full model of ig with canonical link
```

```
#b
newdata<-data.frame(Brand = "Coke", Formulation = "Diet")
pred<-predict(ig.full, newdata=newdata, type="response")
pred
```

```
##       1
## 46.3602
```

```
xf<-c(1,0,0,0)
Var.Yf<-summary(ig.full)$dispersion
D.f<- pred^3/(-2)
Var.ef<-Var.Yf+(D.f^2)*t(xf)%*%vcov(ig.full)%*%xf
Var.ef
```

```
##           [,1]
## [1,] 0.6348606
```

```
lower.yf<-pred-qnorm(0.9)*sqrt(Var.ef)
upper.yf<-pred+qnorm(0.9)*sqrt(Var.ef)

lower.yf
```

```
##          [,1]
## [1,] 45.33908
```

```
upper.yf
```

```
##          [,1]
## [1,] 47.38132
```

```
#c
lrtest(ig.2, ig.main)$"Pr(>Chisq)"[2]
```

```
## [1] 1.039708e-13
```

```
lrtest(ig.2, ig.full)$"Pr(>Chisq)"[2]
```

```
## [1] 4.002091e-21
```

```
# So, brand is statistically significant
```